Computational Scientometrics: Theory and Applications

Cornelia Caragea
Department of Computer Science and Engineering
University of North Texas, USA
ccaragea@unt.edu

C. Lee Giles
Information Science and Technologies
Pennsylvania State University, USA
giles@ist.psu.edu

Vetle I. Torvik
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign, USA
vtorvik@illinois.edu

Lior Rokach
Department of Information Systems Engineering
Ben-Gurion University of the Negev, Israel
lior.rokach@gmail.com

Keywords: bibliometrics, digital libraries, citation analysis, mining citation graphs

---

The field of Scientometrics is concerned with the analysis of science and scientific research. As science advances, scientists around the world continue to produce large numbers of research articles, which provide the technological basis for worldwide collection, sharing, and dissemination of scientific discoveries. Research ideas are generally developed based on high quality citations. Understanding how research ideas emerge, evolve, or disappear as a topic, what is a good measure of quality of published works, what are the most promising areas of research, how authors connect and influence each other, who are the experts in a field, what works are similar, and who funds a particular research topic are some of the major foci of the rapidly emerging field of Scientometrics.

Digital libraries and other databases that store research articles have become a medium for answering such questions. *Citation analysis* is used to mine large publication graphs in order to extract patterns in the data (e.g., citations per article) that can help measure the quality of a journal. *Scientometrics*, on the other hand, is used to mine graphs that link together multiple types of entities: authors, publications, conference venues, journals, institutions, etc., in order to assess the quality of science and answer complex questions such as those listed above. Tools such as *maps of science* that are built from digital libraries, allow different categories of users to satisfy various needs, e.g., help researchers to easily access research results, identify relevant funding opportunities, and find collaborators. Moreover, the recent developments in data mining, machine learning, natural language processing, and information retrieval makes it possible to transform the way we analyze research publications, funded proposals, patents, etc., on a web-wide scale.

The workshop aims at bringing together researchers with diverse interdisciplinary backgrounds interested in mining large digital libraries and other relevant databases. The topics of interest include, but are not limited to:

- New approaches to measuring the impact of research publications as well as the impact of researchers in a particular field of study.
    - Identifying influential authors, experts, and collaborators within or across disciplines
    - Modeling the referencing behavior across disciplines
    - Automatic citation recommendation

Caragea, C., Giles, C. L., Torvik, V. I., & Rokach, L. (2013). Computational Scientometrics: Theory and applications. *iConference 2013 Proceedings* (pp. 1017-1018). doi:10.9776/13429

- Mining large digital libraries of scientific publications and linking to other databases such as funded proposals and patents:
    - Identifying research trends and topics
    - Extracting relevant information from research articles
    - Scaling up machine learning algorithms to large datasets
    - Classification and clustering of scientific publications, funded proposal, patents, etc.
    - Linking together various entities, e.g., articles with articles by similarity, articles with their corresponding presentation slides, articles with the corresponding funded proposals.

- Presenting open-access, novel datasets (e.g., based on Wikipedia, DBpedia, United States Census Bureau data) that can be linked to entities in digital libraries, and can help researcher develop novel technologies for analyzing scientific publications.